

# Crystal structure of 2'-O-Me(CGCGCG)<sub>2</sub>, an RNA duplex at 1.30 Å resolution. Hydration pattern of 2'-O-methylated RNA

Dorota A. Adamiak\*, Jan Milecki<sup>1</sup>, Mariusz Popena, Ryszard W. Adamiak, Zbigniew Dauter<sup>2</sup> and Wojciech R. Rypniewski<sup>2</sup>

Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland, <sup>1</sup>Faculty of Chemistry, Adam Mickiewicz University, Poznan, Poland and <sup>2</sup>European Molecular Biology Laboratory, c/o DESY, Hamburg, Germany

Received July 14, 1997; Revised and Accepted October 6, 1997

NDB coordinates ARFS26

## ABSTRACT

The molecular and crystal structure of 2'-O-Me(CGCGCG)<sub>2</sub> has been determined using synchrotron radiation at near-atomic resolution (1.30 Å), the highest resolution to date in the RNA field. The crystal structure is a half-turn A-type helix with some helical parameters deviating from canonical A-RNA, such as low base pair rise, elevated helical twist and inclination angles. In CG steps, inter-strand guanines are parallel while cytosines are not parallel. In steps GC this motif is reversed. This type of regularity is not seen in other RNA crystal structures. The structure includes 44 water molecules and two hydrated Mg<sup>2+</sup> ions one of which lies exactly on the crystallographic 2-fold axis. There are distinct patterns of hydration in the major and the minor grooves. The major groove is stabilised by water clusters consisting of fused five- and six-membered rings. Minor groove contains only a single row of water molecules; each water bridges either two self-parallel cytosines or two self-parallel guanines by a pair of hydrogen bonds. The structure provides the first view of the hydration scheme of 2'-O-methylated RNA duplex.

## INTRODUCTION

RNA forms a wide range of functionally important tertiary structural domains containing both single- and double-stranded regions such as hairpins, bulge loop duplexes, pseudoknots or hammerheads. The tendency to form double-helical regions plays a crucial role in the folding of these complex structures. Double-stranded RNA helices exist principally in the right-handed A-form. Averaged helical parameters for this canonical form of RNA have been deduced from fibre diffraction data (1). Since then, single crystal analysis of dinucleotide monophosphates (2,3) and refined tRNA structures (4,5) have been reported showing more details of local effects of base sequence. Thanks to recent advances in chemical (6) and enzymatic (7) oligoribonucleotide synthesis

the number of X-ray RNA duplex structures is increasing. However, to our knowledge, only a dozen duplex structures have been reported to date. The first was r[U(UA)<sub>6</sub>A]<sub>2</sub>, at 2.25 Å resolution, with two unexpected kinks in the helix (8). Other duplex structures (9–16), usually form approximately one-turn helices. The resolution of these structures, some of which contain more than one mismatch, varies from 1.6 to 2.6 Å. Most recently, structure of a 160 nt long domain of a class I intron at 2.8 Å (17) and the hammerhead ribozymes (18,19) have been reported.

Our continued interest in RNA duplexes containing alternative CG base pairs was prompted by the finding that poly[r(CG)] (20) and r(CGCGCG)<sub>2</sub> (21) are able to form left-handed RNA double helices named Z-RNA. This intrinsic property of (CG)<sub>n</sub> duplexes is interesting in view of the structural properties of RNA and RNA hydration. The latter phenomenon is known to govern helicity reversal processes (20,21). Recently, it was found that the A→Z transition of (CG)<sub>n</sub> duplexes is also promoted by high pressure (22). In contrast, 2'-O-Me(CGCGCG)<sub>2</sub> does not undergo helicity reversal under similar experimental conditions and remains in the A-form (23). This points to the influence of 2'-O-methylation on the ability of RNA duplexes, containing alternating CG base pairs, to undergo such a conformational reversal.

To date, there is no X-ray RNA structure containing alternating CG base pairs which would allow analysis of both CG and GC steps. Two high resolution structures, r(UUCGCG)<sub>2</sub> (24) and r(CCCCGGGG)<sub>2</sub> (25) solved at 1.40 and 1.46 Å, respectively, have been reported recently. Although they contain CG base pairs, the duplex structures formed do not allow an analysis of both CG and GC steps. The r(CCCCGGGG)<sub>2</sub> (25) has only one CG step and no GC steps. In the crystal of UUCGCG hexamer, a duplex structure with four CG base pairs and two overhanging UU base pairs is formed (24). Only the GC step lies in the interior of the structure while the two CG steps contain terminal guanosine residues. Due to crystal symmetry relating the two UUCGCG strands the structure contains only one CG step and one GC step with a crystallographic 2-fold axis lying in the middle of the GC step which means that only one half of the GC step is unique.

\*To whom correspondence should be addressed. Tel: +48 61 8 528503; Fax: +48 61 8 520532; Email: dorota@tryptophan.ibch.poznan.pl

Unfortunately, our extensive attempts to get suitably diffracting monocrystals of r(CGCGCG)<sub>2</sub> duplex have been unsuccessful.

In this work we present the crystal structure of the 2'-O-Me(CGCGCG)<sub>2</sub> duplex at 1.30 Å resolution. The near-atomic resolution of the X-ray data allowed very accurate structure determination with anisotropic thermal parameters and detailed analysis including the hydration scheme and magnesium binding. The overall structure is an A-RNA helix, but with certain structural features which deviate from the canonical form. The results obtained also provide an insight into the effect of 2'-O-methylation on hydration of RNA duplex.

In the accompanying paper we described the NMR structure of r(CGCGCG)<sub>2</sub> and 2'-O-Me(CGCGCG)<sub>2</sub> under low salt conditions. Surprisingly, the two right-handed duplex structures are similar, despite 2'-O-methylation, with an average r.m.s.d. of 1.0 Å. This suggests that it is the intrinsic properties of alternating CG base pairs that govern both RNA duplex structures. The data allow for comparison of the structure of a 2'-O-methylated RNA duplex in the crystalline state and solution. We hope that the results, for the 2'-O-methylated analogue, will bring us closer to the understanding of the structure of native (CG)<sub>n</sub> sequences.

## MATERIALS AND METHODS

### Oligoribonucleotide crystals

Hexamer 2'-O-Me(CGCGCG) was prepared by automated solid-phase synthesis using phosphoramidite chemistry (26). Duplex crystals were grown at 20°C by hanging drop/vapour diffusion. After an extensive search several monocrystals were obtained under the following conditions: 5 mg/ml of RNA in 50 mM HEPES buffer pH 7.5, 15 mM MgCl<sub>2</sub>, 1 mM spermine tetrahydrochloride and 30–40% 2-methyl-2,4-pentanediol as precipitating agent. No crystals of the respective octamer could be obtained under these or similar conditions.

### Crystallographic data collection and processing

X-ray diffraction data were collected from a single crystal on the EMBL BW7B wiggler beam line at the DORIS storage ring, DESY, Hamburg, with a Mar Research imaging plate scanner. The crystal was mounted with the long *c*-axis along the goniometer spindle axis. Due to the highly elongated unit cell this orientation was essential to avoid overlaps of diffraction intensities. Three data sets were collected: at long, medium and short exposures, to record intensities at high, medium and low resolution. The intensities were integrated using the program DENZO and scaled using program SCALEPACK (27). Outliers were rejected based on the chi-square test implemented in SCALEPACK. The post-refinement option was used to refine the cell parameters. The X-ray data are summarised in Table 1.

### Structure solution and refinement

The solvent content of the crystal was calculated to be 47% assuming RNA density 1.7 g/cm<sup>3</sup> (28) and one duplex per asymmetric unit. The structure was solved by molecular replacement as implemented in the program AMORE (29) from the CCP4 program suite (30). The solution was obtained using residues U6 to A11 of chain A and U4 to A9 of chain B from the

U(UA)<sub>6</sub>A duplex RNA structure (9,10, PDB code 1RNA) as the starting model, which corresponded to the (UA)<sub>3</sub> core of the duplex. No further editing was done on the starting model at this stage. The rotation function was calculated using terms between 8 and 2.5 Å, with a Patterson search radius of 12 Å. The 50 highest peaks of the rotation function did not show any clear candidates for the correct solution. The correlation coefficient decreased smoothly from 0.40 to 0.34 for the first 19 peaks. Further peaks had a correlation coefficient of ~0.2. The translation function was calculated for each of the first 20 peaks in the rotation function. It was not known at this stage if the space group was P6<sub>1</sub>22 or P6<sub>5</sub>22 and the translation function was calculated for both space groups. The best solution was obtained for P6<sub>1</sub>22, with correlation coefficient 0.67. The second highest peak had correlation coefficient 0.54 and the average value was ~0.5. After rigid body refinement the correlation coefficient for the highest peak was 0.74, for the second highest peak 0.58 and the average value was still ~0.5. The model was positioned in the unit cell according to the highest peak and (3F<sub>o</sub> - 2F<sub>c</sub>) and (F<sub>o</sub> - F<sub>c</sub>) difference maps were inspected. It became clear that this was the correct solution, with the (3F<sub>o</sub> - 2F<sub>c</sub>) map showing good overall agreement with the model although considerable deviations were observed for the terminal base pairs. The difference map (F<sub>o</sub> - F<sub>c</sub>) showed most of the 2'-O-methyl groups and exo-amino groups of C and G, not included in the model. There were no bad intermolecular contacts.

The structure was refined by stereochemically restrained least-squares minimisation as implemented in the program SHELXL93 (31). The integrated diffraction intensities between 8 and 1.30 Å were used in the refinement, rather than the structure factor amplitudes. The geometric restraints were derived from the standard dictionary used in the CCP4 program suite (30). Planarity restraints were imposed on the guanine and the cytosine rings, as well as restraints on bond lengths and bond angles. In the later stages of refinement the bond angle restraints were removed. The initial model was modified to reflect the correct chemistry and the non-hydrogen atoms that were initially absent in the model were easily found in the electron density maps. Cycles of least-squares refinement were interspersed with rounds of manual rebuilding, based on (3F<sub>o</sub> - 2F<sub>c</sub>) and (F<sub>o</sub> - F<sub>c</sub>) maps, using an Evans and Sutherland ESV graphics station and the program FRODO (32). Initially only non-hydrogen atoms were included in the model and isotropic temperature factors were refined. When hydrogen atoms became visible in the (F<sub>o</sub> - F<sub>c</sub>) map they were included in the model and anisotropic temperature factors were refined. Solvent molecules were inserted manually, based on examination of electron density maps. Refinement was terminated when it was felt that no further significant improvement in the model could be achieved. The final *R*-factor (Σ||F<sub>o</sub>| - |F<sub>c</sub>||/Σ|F<sub>o</sub>|) was 0.175. The refinement was performed using the conjugate gradient algorithm. After the refinement was completed, one additional cycle of minimisation was executed using the full-matrix least-squares method in order to obtain direct estimates of errors in atomic positions and temperature factors. To decrease the size of the computation, the model was divided into three blocks, two containing one RNA strand each and one block for the solvent atoms. All restraints and shift damping were removed in that cycle. Helical parameters were calculated (33) using program CURVES 5.11. The SGI Indigo<sup>2</sup> workstation was used for visualisation applying the InsightII/Biopolymer software package (MSI).

**Table 1.** Summary of X-ray data

	High	Medium	Low	Total
Number of images	35	12	36	82
Oscillation angle	1°	3°	5°	
Maximum resolution	1.30 Å			
Wavelength	0.857 Å			
% $R_{\text{merge}}^a$	7.0			
Raw measurements used	96 765			
Unique reflections	8165			
% completeness (25–1.3 Å)	95.3			
% $>3\sigma$	80.2 (48.1 in high res. bin 1.32–1.30 Å)			
Space group	$P6_122$			
Postrefined cell param. (Å)				
$a = b$	26.27			
$c$	160.90			

<sup>a</sup> $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum \langle I \rangle$ , where  $I_i$  is an individual intensity measurement, and  $\langle I \rangle$  is the average intensity for this reflection with summation over all the data.

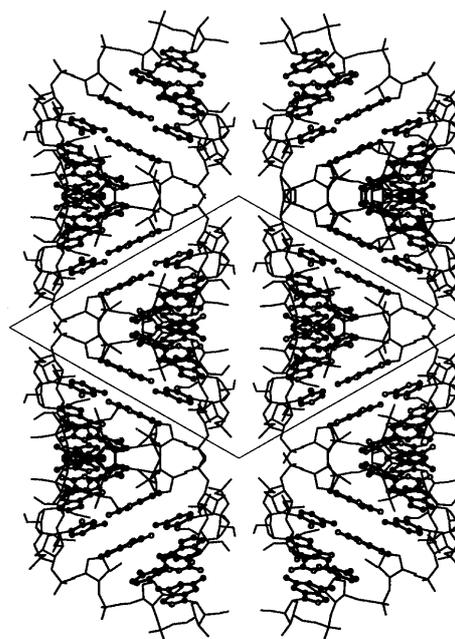
## RESULTS AND DISCUSSION

### The refined structure

The high quality of the crystals and the use of synchrotron radiation enabled refinement of the model of the 2'-*O*-Me(CGCGCG)<sub>2</sub> duplex structure with an exceptional accuracy. Anisotropic thermal vibrational parameters and direct estimates of standard deviations for atomic positions and the temperature factors have been obtained. A measure of reliability of refinement is gained from examining the data-to-parameter ratio (*d/p*). For the refined model of 2'-*O*-Me(CGCGCG)<sub>2</sub> the final value of *d/p* was 2.9, with anisotropic temperature factors. When restraints are taken into account the *d/p* value becomes effectively higher. Thus the refinement process is well determined even with anisotropic temperature factors. The model of the 2'-*O*-methylated RNA duplex is complete. It also includes 44 solvent water molecules and two Mg<sup>2+</sup> ions, one of which lies exactly on the 2-fold crystallographic axis. The second magnesium site is only partially occupied but is clearly recognisable by its octahedral coordination. No other ordered sites were found for Mg<sup>2+</sup> or spermine, which were both present and necessary for crystallisation.

**Crystal packing.** The asymmetric unit contains one duplex. The crystal lattice consists of infinite columns of hexamer double helices stacked head-to-tail, perpendicular to the *c*-axis. The columns are arranged in layers, with a 60° rotation between layers. Two stacked hexamers form one full turn of a helix (Fig. 1). The requirement of crystallographic symmetry means that there are exactly 12 bp per turn of the column.

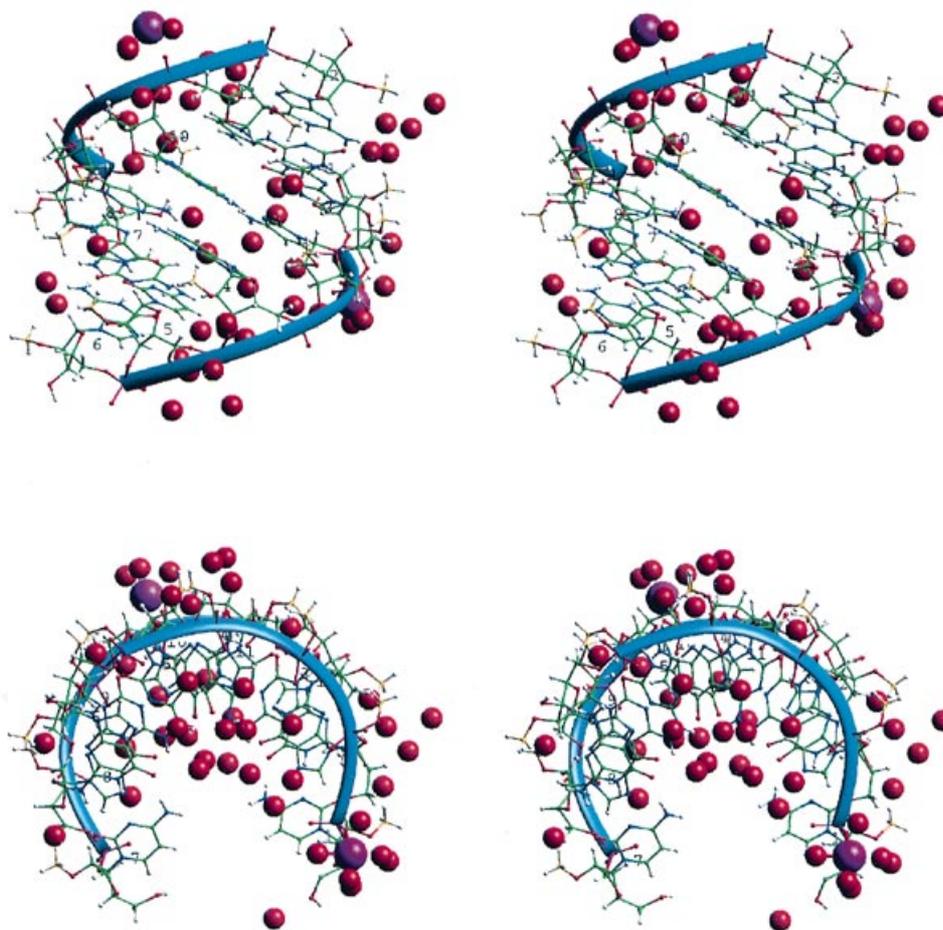
**Accuracy of the coordinates.** The overall estimated standard deviation (e.s.d.) for atomic positions is 0.090 Å. For the RNA atoms it is 0.077 Å (oxygen 0.063 Å, carbon 0.091 Å, nitrogen 0.062 Å, phosphorus 0.033 Å) and for the water molecules 0.146 Å.



**Figure 1.** Crystal packing of the 2'-*O*-Me(CGCGCG)<sub>2</sub>. The crystal lattice consists of infinite columns of double helices stacked head-to-tail, perpendicular to the *c*-axis. The columns are arranged in layers, with a 60° rotation between layers.

### Structural features of 2'-*O*-Me(CGCGCG)<sub>2</sub>

**The overall structure.** In the crystal, self-complementary hexamer 2'-*O*-Me(CGCGCG) forms approximately one half-turn of an RNA helix (Fig. 2). Both strands are related by non-crystallographic 2-fold symmetry axis with r.m.s.d. values of 0.19 Å. The overall structure is an A-type RNA helix but with certain deviations from the canonical, fibre RNA structure (1). The deviation is much lower in



**Figure 2.** Stereo view of the molecular structure of the 2'-O-Me(CGCGCG)<sub>2</sub> duplex showing 44 water molecules (red spheres) and two hydrated magnesium ions (violet spheres). The position of the backbone of the two strands (C1–G6 and C11–G16) is indicated by ribbons.

the crystal structure (r.m.s.d. 1.30 Å) than in solution (r.m.s.d. 1.8 Å). The duplex is overwound and contains 10 bp per helical turn. The difference from the number of base pairs per turn in the crystal packing is due to dislocation in the intermolecular helix stacking (co-axial stacking). The NMR structures of r(CGCGCG)<sub>2</sub> and 2'-O-Me(CGCGCG)<sub>2</sub> described in the accompanying paper are closely similar. The r.m.s.d. between the 2'-O-Me(CGCGCG)<sub>2</sub> X-ray structure and the solution structure is 1.7 Å. The possible causes are the influence of crystal packing and differences in refinement methods and restraints.

**2'-O-methylribose, glycosidic bond and backbone conformation.** The non-equivalence of both strands is reflected by differences in the sugar pucker (34) and backbone conformational parameters. All sugars are in stabilised C3'-endo pucker with average  $P_N$  value of 13°. Residues G2 and G12 show C2'-exo ( $P = -3^\circ$ ) and C3'-endo ( $P = 34^\circ$ ) pucker, respectively. All residues are characterised by high pucker amplitudes (average  $\Phi = 39^\circ$ ) and by *anti* glycosidic bond angles. The latter and the  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$  backbone torsion angles, (Table 2) fall in the range typical for the A-family of right handed helices (35). The  $\gamma$  torsion angle ranges from 41 to 68° describing their (+) *gauche* conformations. The values correspond to those obtained for 2'-O-Me(CGCGCG)<sub>2</sub> in solution by coupling constants analysis but they are ~20–25° lower than the values given by NMR

structure refinement. Anticorrelated  $\alpha$  torsion angles are proportionally lower. The  $\beta$  (178–183°) and  $\epsilon$  (201–209°) torsion angles indicate a favoured *trans* conformation of corresponding bonds.

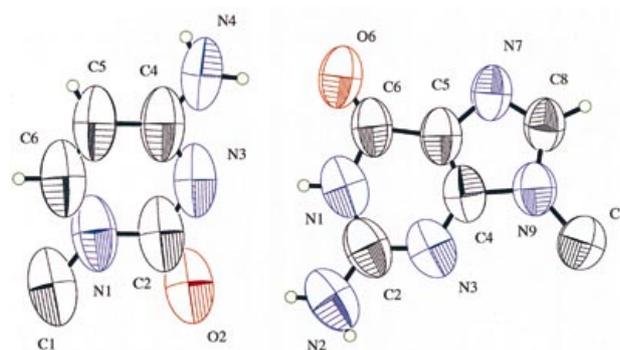
**Base pairs and stacking geometry.** Watson–Crick base pairing is observed throughout the duplex. Anisotropic temperature factors point to significant lateral disorder of the bases, most pronounced in the direction perpendicular to the inter-base hydrogen bonds (Fig. 3).

Analysis of helical parameters (33) (Table 3), reveals that the rise parameter within the core of the duplex is small (2.2 Å), there is no alternation for CG and GC steps and its value rises substantially at the ends of the duplex. The x-displacement parameter is in the range of the canonical structure and much higher than observed in solution. Helical twist angles are higher than typical for A-RNA helices. The inclination and propeller twist angles are elevated but to a smaller extent than in solution. A characteristic stacking pattern was observed for the 2'-O-Me(CGCGCG)<sub>2</sub> structure; much more regular than seen in solution (Fig. 4). Within CG steps, inter-strand guanines are parallel while cytosines are not parallel. In GC steps this motif is exactly reversed. This motif is not seen in other RNA crystal structures. An analysis of the (UA)<sub>6</sub> fragment within crystal RNA duplex structure (9,10) reveals that all inter-strand adenines are

nearly parallel pairwise but uracils are not. In 2'-O-Me (CGCGCG)<sub>2</sub>, inter-strand, parallel cytosines and guanines interact with well defined water molecules in the minor groove (see below). In CG steps considerable inter-strand stacking of guanines is observed (Fig. 4). Such a pattern of purine inter-strand stacking in pyrimidine-purine steps is a typical feature of A-RNA helices (35). In contrast, inter-strand guanine stacking in CG steps in the solution structure is very limited. Values of the roll angles are rather low and in the core of the duplex show no tendency to alternate as observed for the structure in solution. Interestingly, apart from terminal base pairs, the roll angles are positive for both steps indicating their tendency to be open toward the minor groove.

### Magnesium binding

Only two different hydrated magnesium sites were located in the crystal lattice, at half occupancy each. Due to the high quality and resolution of the X-ray data it is unlikely that additional ordered magnesium sites exist unobserved. They most probably are delocalised, forming dispersed cation shield within the structure. One magnesium cation lies exactly on the crystallographic 2-fold axis and bridges two symmetry related molecules through two C5 phosphate oxygens (O2P) and four water molecules (Fig. 5). The 2'-hydroxyl function has been found in several crystal structures to play a crucial role in RNA-RNA water-mediated intermolecular contacts (10,16,24,25). However, in this structure 2'-OH groups are methylated and blocked as H-bond donors. Instead, the magnesium cation and the two 2'-oxygens play the pivotal role in



**Figure 3.** Thermal-ellipsoid, at 25% probability, representation of the Watson-Crick base pair C1 (left) and G16 (right) in the 2'-O-Me(CGCGCG)<sub>2</sub> duplex. Anisotropic temperature factors point to significant lateral disorder of the bases.

bridging symmetry related molecules in the lateral, column-to-column direction. The other magnesium site is not fully occupied in the crystal lattice and was refined at an occupancy factor 0.5. It is coordinated by one of the oxygens of the G2 phosphate and five water molecules; one of them is in close contact of 2.69 Å (H-bond?) to a symmetry related G4 O1P. Both magnesium cations are located near duplex ends and it is possible that they contribute to the 'edge effects' seen in the helical parameters, rise and roll, for the terminal base steps (Table 3).

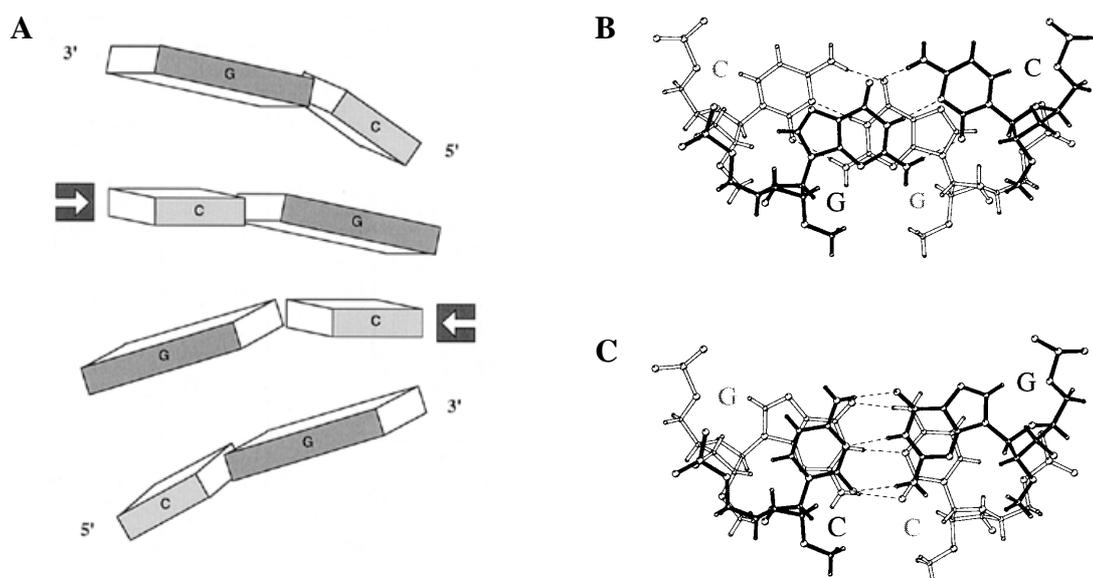
**Table 2.** Sugar, backbone and glycosidic torsion angles<sup>a</sup> for 2'-O-Me(CGCGCG)<sub>2</sub> structure from the X-ray study (left column). The results from the NMR refinement<sup>b</sup> (see accompanying paper) are quoted for comparison (right column)

Residue	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\chi$								
1st strand	2nd strand														
C1				41	72(9)	93	86(4)	202	197(4)	288	285(4)	202	201(5)		
	C11			52		76		219		280		193			
G2		297	277(9)	176	180(4)	46	70(7)	82	87(3)	215	203(4)	291	291(5)	193	202(4)
	G12			177		51		93		203		287		195	
C3		277	267(7)	172	177(5)	61	75(5)	68	88(4)	219	201(4)	290	288(5)	197	213(5)
	C13			170		49		72		217		290		198	
G4		288	262(8)	177	179(5)	55	79(6)	75	93(3)	215	208(4)	285	294(4)	200	207(5)
	G14			176		54		76		215		290		198	
C5		293	247(8)	173	177(5)	54	85(6)	72	82(3)	207	201(4)	296	288(4)	196	206(5)
	C15			171		52		73		211		291		197	
G6		291	274(8)	182	173(4)	63	72(6)	75	82(4)	-	-	-	-	201	191(5)
	G16			174		63		72		-	-	-	-	199	
Mean		291	265	175	177	53	75	77	86	212	202	288	289	198	203
A-RNA <sup>c</sup>		186		49		95			202		294		202		

<sup>a</sup> $\alpha$  O5'  $\beta$  C5'  $\gamma$  C4'  $\delta$  C3'  $\epsilon$  O3'  $\zeta$  P.

<sup>b</sup>Standard deviations are given in parentheses.

<sup>c</sup>Refs 1,35,36.



**Figure 4.** The pattern of inter-strand stacking within CG (B) and GC steps (C) of the 2'-O-Me(CGCGCG)<sub>2</sub> duplex. In CG steps, inter-strand guanines are parallel while cytosines are not parallel; in steps GC this motif is reversed (A). One pair of parallel cytosines is indicated by arrows. Two pairs of parallel guanines are also shown. This pattern is observed for all the duplex, although some bases have been omitted from this diagram for clarity.

**Table 3.** Helical parameters<sup>a</sup> for 2'-O-Me(CGCGCG)<sub>2</sub> structure as resulted from X-ray (left column) and NMR (right column) studies

Base pairs	x-Displacement (Å)		Inclination (°)		Propeller twist (°)	
C1-G12	-4.8	-3.4	19	23	-6	-5
G2-C11	-4.8	-3.2	19	26	-16	-17
C3-G10	-4.9	-3.3	21	28	-12	-26
G4-C9	-5.0	-3.2	22	30	-8	-38
C5-G8	-5.0	-3.3	21	24	-6	-25
G6-C7	-5.0	-3.1	22	20	-7	-23
Base steps	Rise (Å)		Twist (°)		Roll (°)	
C1-G2	3.2	2.3	40	34	-4	0
G2-C3	2.2	2.4	34	37	6	-3
C3-G4	2.2	2.6	34	42	4	6
G4-C5	2.2	2.6	38	41	4	-6
C5-G6	2.9	2.7	37	34	-4	1

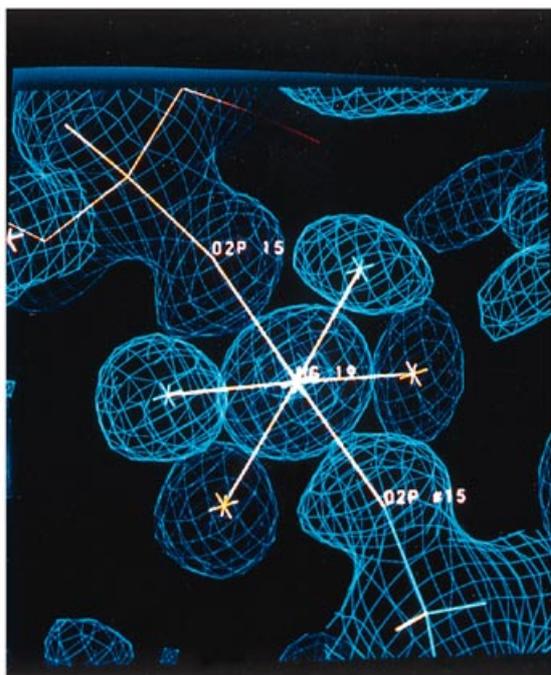
<sup>a</sup>Parameters were calculated (33) with program CURVES 5.11.

### Hydration pattern of 2'-O-methylated RNA duplex

The major groove of 2'-O-Me(CGCGCG)<sub>2</sub> is narrow but deep and the minor groove is broad and shallow. The 2'-O-methyl groups point towards the minor groove (Fig. 2). The duplex grooves are hydrated in a very regular way, with the majority of ordered waters located in the major groove. Although no hydrogen atoms were observed directly for the water molecules, the position of the hydrogens could be deduced for many waters in the first hydration shell within ~3.5 Å from RNA sites (Fig. 6). This was possible because of the special regular arrangement of the waters and the chemical nature of the groups hydrated. The

pattern of hydration in this structure can be compared to the related crystal structure of r(CCCCGGGG)<sub>2</sub> at 1.46 Å resolution; the only relevant case for which the RNA duplex hydration scheme has been presented in detail (25).

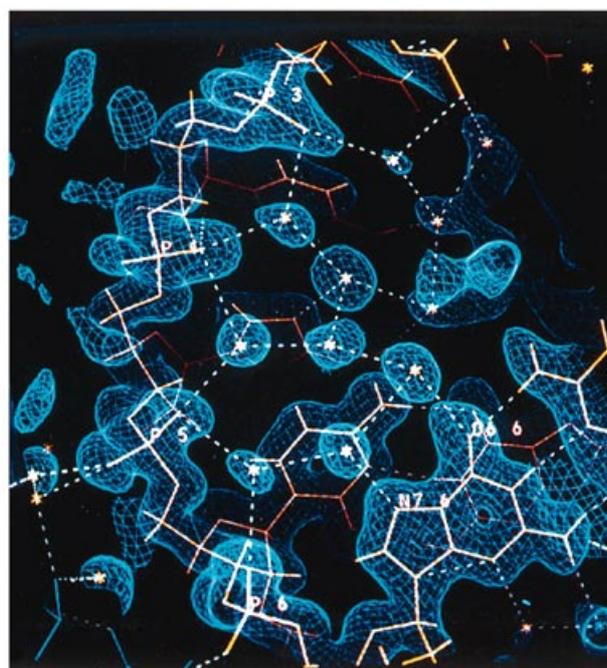
*Major groove hydration.* The water molecules in the major groove form a regular repetitive pattern. As in other A-DNA (37-39) and A-RNA (24,25) oligonucleotide crystal structures, the intra-strand phosphate oxygens (O1P) are water bridged with W-O1P distances ranging from 2.8 to 3.3 Å. Those bridges form one edge of a set of water clusters consisting of alternating, fused five- and six-membered rings (Fig. 7). The six-membered rings



**Figure 5.** The hydrated magnesium cation at the C15 (O2P) site; a region of the  $(3F_0 - 2F_c)$  electron density map.

are nearly planar and the five-membered rings are non-planar. The pattern is observed twice, once for each RNA strand. Stable, low energy assemblies of water molecules have been studied extensively (40). Low energy pentagons are often found in nucleic acid duplexes (24,25,35,37,41). To our knowledge, formation of fused, nearly planar hexameric rings composed of five waters and phosphate oxygen is not seen in other RNA crystal structures (Fig. 8). The network is completed by H-bonds (average distance 2.8–2.9 Å) with N<sup>7</sup> and O<sup>6</sup> of guanines as primary base sites. The N<sup>4</sup>-H of cytosines are less strongly involved with an average distance of 3.3 Å (Fig. 7).

One difficulty with the H-bond network in the major groove (Fig. 7), is that there are more observed hydrogen bonds than there are hydrogens in the structure, assuming canonical protonation. One possibility is that there is a net deficit of protons but the protons present are shared, perhaps by means of a fast exchange mechanism. Although this is not directly observed in this structure, bifurcated H-bonds, with a single hydrogen shared between two acceptors, are often observed in atomic resolution X-ray studies. Another possibility is that there is an additional proton within the network. Although the pH of crystallisation, 7.5, is unfavourable for an excess of protons, pH being a bulk phenomenon does not necessarily apply locally. An additional proton within the H-bond network could partly neutralise the negative charge of the phosphate backbone and play a structural role by stabilising the H-bonding network of water and the RNA structure. The details of the H-bonding (Fig. 7) are hypothetical and only schematic. The detailed picture is likely to be more complex. An additional complication is the presence of Mg<sup>2+</sup> ions near the phosphate groups. Whatever the underlying proton structure, the overall observed hydration pattern is clear (Fig. 7, left). Its regularity even within the relatively short sequence of the

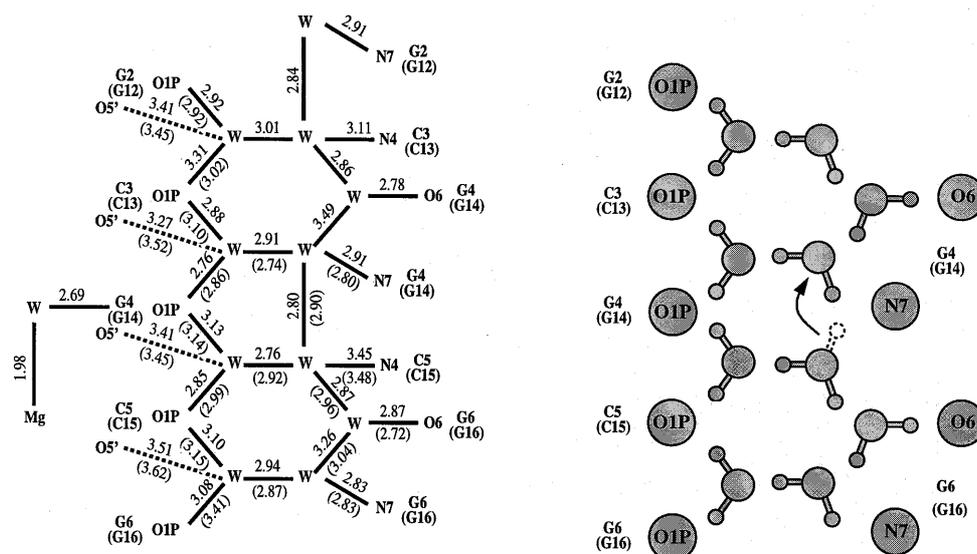


**Figure 6.** Hydration of the 2'-O-Me(CGCGCG)<sub>2</sub> duplex within the major groove; a region of the  $(3F_0 - 2F_c)$  electron density map.

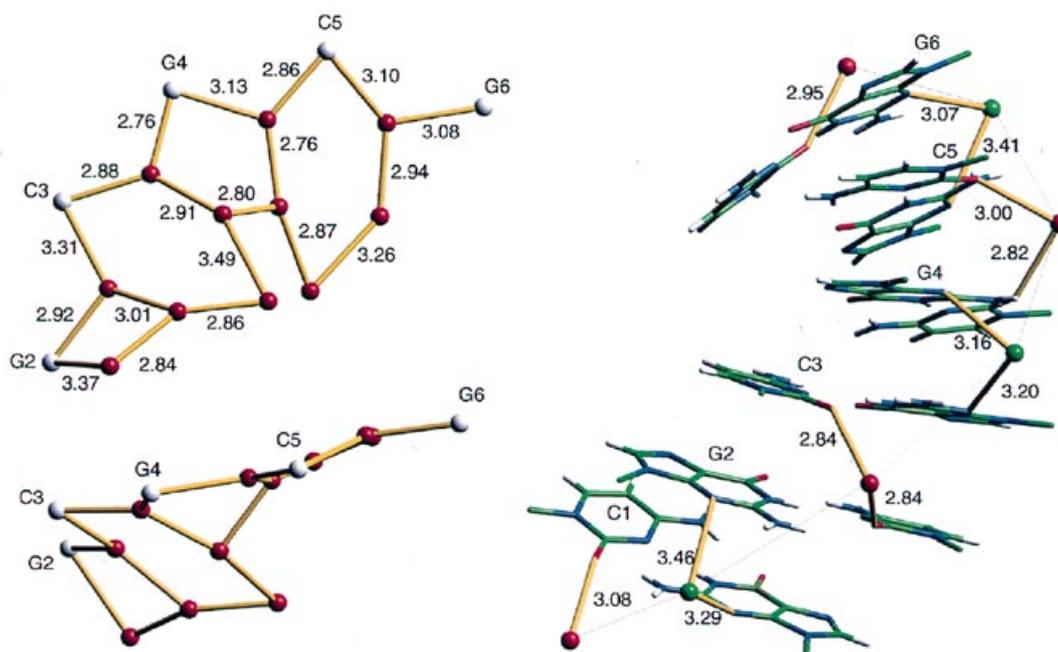
hexamer duplex implies that the bound water molecules play an important role in stabilising the molecule.

**Minor groove hydration.** The minor groove is hydrated by a single row of water molecules (Fig. 8). Waters in the minor groove are not in contact with each other (average distance 6.3 Å) and span both oligoribonucleotide strands. They act as strong H-bond donors. Each water bridges by a pair of hydrogen bonds, either two self-parallel cytosines or two self-parallel guanines. Waters hydrogen bonded to the O<sup>2</sup> of cytosines are at distances 2.8–3.1 Å. Waters bridging the endocyclic N<sup>3</sup> of guanine are bound less strongly (3.1–3.5 Å). In addition, all the waters are weak acceptors from exo-NH<sub>2</sub> groups of the guanines with average H-bond distances 3.6 Å. This is in accord with NMR observations on fast water exchange with guanine exo-NH<sub>2</sub> groups of 2'-O-Me(CGCGCG)<sub>2</sub> (see accompanying paper). All the waters are clearly defined in the  $(3F_0 - 2F_c)$  map with average temperature factors in the range 25–32 Å<sup>2</sup>. The hydration pattern in the minor groove extends across the axially stacked RNA molecules, with single water sites, lying exactly on the 2-fold axes, bridging C1 and C11 with their symmetry-related equivalents. Whereas all the other water-bridged bases, both cytosines and guanines, are parallel within the same molecule, the two pairs of cytosines water-bridged between the stacked molecules are not parallel. The pattern of water-mediated inter-base contacts in the minor groove offers an explanation of how the bases maintain their distinctive cross-strand parallel arrangement.

With no crystal structure of r(CGCGCG)<sub>2</sub> available, a discussion of the effect of 2'-O-methylation on the RNA minor groove hydration cannot be complete. In the crystal structure of r(CCCCGGGG)<sub>2</sub> (25), the minor groove is hydrated in a complex way with two rows of water molecules across the groove, spanning



**Figure 7.** Hydration of the 2'-O-Me(CGCGCG)<sub>2</sub> duplex in the major groove. The pattern of alternating, fused five- and six-membered rings can be seen for both RNA strands (left). H-bonding donor-acceptor distances are shown in Å for strand C1-G6 and strand C11-G16 (values in brackets) for the strand C11-G16 are disordered, therefore, no distances are given. Proximity of the OIP sites to the O5' is marked by dotted lines. (Right) Schematic showing deduced orientation of the water molecules. A hypothetical additional proton is drawn on one of the water molecules to illustrate that there are less protons in the structure than there are hydrogen bonds in this alternating five- and six-membered ring H-bonded network.



**Figure 8.** The water shell architecture within the 2'-O-Me(CGCGCG)<sub>2</sub>. The extensive hydration of the intra-strand phosphates (oxygens in white) in the major groove by water clusters (solid yellow lines and water oxygens in red) consisting of fused, nearly planar six-membered rings (one of the rings, in lower part of the figure, is shown edge-on) and five-membered rings (left) versus simple, rigid hydration within the minor groove (right). The single row of waters stabilises inter-strand base-base interactions. Each water bridges by a pair of hydrogen bonds either two self-parallel cytosines (waters in red) or two self-parallel guanines (waters in green).

free 2'-OH groups and maintaining hydrogen bonding with O<sup>2</sup> of the cytosine and N<sup>2</sup> and N<sup>3</sup> of guanine residues. In 2'-O-Me(CGCGCG)<sub>2</sub>, the 2'-O-methyl groups point towards the minor groove leading to its narrowing and increasing the hydrophobicity of the region (Fig. 2). None of the waters within the minor groove have access to 2'-oxygens as H-bond acceptors. Such a pattern of

hydration influences non-bonded hydrophobic interactions between inter-strand 2'-O-methyl groups. The average distance between inter-strand 2'-O-methyl carbon atoms is 7.7 Å (r.m.s. 0.1 Å) and is very similar to the distance estimated from the NMR analysis in solution (7.5 Å, r.m.s. 0.4 Å). This means that the distance between 2'-O-methyl groups across the minor groove is only

3.7 Å, after taking account of the van der Waals radii for the two methyl groups. The latter value is lower than the value found by the X-ray analysis of an A-DNA self-complementary duplex containing single 2'-O-methyladenosine residue (4.8 Å) (42). Each water is locked between two inter-strand 2'-O-methyl groups at a distance of ~3.0–3.4 Å between the water oxygen and the nearest methyl hydrogen.

The regularity and clarity of the hydration pattern in both grooves strongly suggests that the bound water molecules play a structural role and should be considered part of the 2'-O-Me (CGCGCG)<sub>2</sub> structure.

## ACKNOWLEDGEMENTS

This work was supported by grants from the State Committee for Scientific Research, Republic of Poland (2P30300704 and 8T11F010008p08). D.A.A. gratefully acknowledges a research fellowship from the Alexander von Humboldt Foundation. The generous support of the Foundation for Polish Science and the Poznan Supercomputing and Networking Centre is acknowledged. Authors thank Prof. Richard Lavery for kindly supplying us with program CURVES 5.11 and Mieczyslaw Kluge for technical assistance.

## REFERENCES

- 1 Arnott, S., Hukins, D.W.L., Dover, S.D., Fuller, W. and Hodgson, A.R. (1973) *J. Mol. Biol.*, **81**, 107–112.
- 2 Seeman, N.C., Rosenberg, J.M., Suddath, F.L., Kim, J.J.P. and Rich, A. (1976) *J. Mol. Biol.*, **104**, 109–144.
- 3 Rosenberg, J.M., Seeman, N.C., Day, R.O. and Rich, A. (1976) *J. Mol. Biol.*, **104**, 145–167.
- 4 Holbrook, S.R., Sussman, J.L., Warrant, R.W. and Kim, S.H. (1978) *J. Mol. Biol.*, **123**, 631–660.
- 5 Hingerty, B., Brown, S.S. and Jack, A. (1978) *J. Mol. Biol.*, **124**, 523–534.
- 6 Usman, N., Ogilvie, K.K., Jiang, M.-J. and Cedergren, R.J. (1987) *J. Am. Chem. Soc.*, **109**, 7845–7854.
- 7 Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) *Nucleic Acids Res.*, **15**, 8783–8798.
- 8 Dock-Bregeon, A.C., Chevrier, B., Podjarny, A., Moras, D., de Bear, J.S., Gough, G.R., Gilham, P.T. and Johnson, J. (1988) *Nature*, **335**, 375–378.
- 9 Holbrook, S.R., Cheong, C., Tinoco, I., Jr and Kim, S.A. (1991) *Nature*, **353**, 579–581.
- 10 Leonard, G.A., McAuley-Hecht, K.E., Ebel, S., Lough, D.M., Brown, T. and Hunter, W.N. (1994) *Structure*, **2**, 483–494.
- 11 Cruse, W., Saludjan, P., Biala, E., Strazewski, P., Prange, T. and Kennard, O. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4160–4164.
- 12 Betzel, C., Lorenz, S., Furste, J.P., Bald, R., Zhang, M., Schneider, T.R., Wilson, K.S. and Erdmann, V.A. (1994) *FEBS Lett.*, **351**, 159–164.
- 13 Baeyens, K.J., De Bondt, H.L. and Holbrook, S.R. (1995) *Nature Struct. Biol.*, **2**, 107–122.
- 14 Schindelin, H., Zhang, M., Bald, R., Furste, J.-P., Erdmann, V.A. and Heinemann, U. (1995) *J. Mol. Biol.*, **249**, 595–603.
- 15 Lietzke, S.E., Barnes, C.L., Berglund, J.A. and Kundrot, C.E. (1996) *Structure*, **4**, 917–930.
- 16 Baeyens, K.J., De Bondt, H.L., Pardi, A. and Holbrook, S.R. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 12851–12855.
- 17 Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996) *Science*, **273**, 1678–1685.
- 18 Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994) *Nature*, **372**, 68–74.
- 19 Scott, W.G., Murray, J.B., Arnold, J.R.P., Stoddard, B.L. and Klug, A. (1996) *Science*, **274**, 2065–2069.
- 20 Hall, K., Cruz, P., Tinoco, I., Jr, Jovin, T.M. and Van de Sande, J.H. (1984) *Nature*, **311**, 584–586.
- 21 Adamiak, R.W., Galat, A. and Skalski, B. (1985) *Biochim. Biophys. Acta*, **825**, 345–352.
- 22 Krzyzaniak, A., Barciszewski, J., Furste, J.P., Bald, R., Erdmann, V.A., Salanski, P. and Jurczak, J. (1994) *Int. J. Biol. Macromol.*, **16**, 159–162.
- 23 Krzyzaniak, A., Salanski, P., Adamiak, R.W., Jurczak, J. and Barciszewski, J. (1996) In Hayashi, R. and Balny, C. (eds), *High Pressure Bioscience and Biotechnology*. Elsevier Science, Amsterdam, pp. 189–194.
- 24 Wahl, M.C., Rao, S.T. and Sundaralingam, M. (1996) *Nature Struct. Biol.*, **3**, 24–31.
- 25 Egli, M., Portman, S. and Usman, N. (1996) *Biochemistry*, **35**, 8489–8494.
- 26 Biala, E., Milecki, J., Kowalewski, A., Popena, M., Antkowiak, W.Z. and Adamiak, R.W. (1993) *Acta Biochim. Pol.*, **40**, 1–7.
- 27 Otwinowski, Z. and Minor, W. (1997) In Carter, C.W. and Sweet, R.M. (eds), *Methods in Enzymology*. Academic Press, New York, Vol. **276**, pp. 307–326.
- 28 Feigin, L.A. and Svergun, D.I. (1987) *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. Plenum Press, New York, p. 120.
- 29 Navaza, J. (1994) *Acta Crystallog. section A*, **50**, 157–163.
- 30 Collaborative Computational Project, Number 4 (1994) *Acta Crystallog. section D*, **50**, 760–763.
- 31 Sheldrick, G.M. and Schneider, T. (1997) In Carter, W.W. and Sweet, R.M. (eds), *Methods in Enzymology*. Academic Press, New York, Vol. **277**, pp. 319–343.
- 32 Jones, T.A. (1978) *J. Appl. Crystallog.* **11**, 268–272.
- 33 Lavery, R. and Sklenar, H. (1988) *J. Biomol. Struct. Dyn.*, **6**, 63–91.
- 34 Altona, C. (1983) *Recl. Trav. Chim. Pays-Bas*, **101**, 413–433.
- 35 Saenger, W. (1984) *Principles of Nucleic Acids Structure*. Springer, Berlin, Heidelberg.
- 36 Chandrasekaran, R. and Arnott, S. (1989) In Saenger, W. (ed.), *Landolt-Bornstein, New Series, Group VII*. Springer, Berlin, Vol. **1b**, pp. 31–170.
- 37 Kennard, O., Cruse, W.B., Nachman, J., Prange, T., Shakked, Z. and Rabinowitch, D. (1986) *J. Biomol. Struct. Dyn.*, **3**, 623–627.
- 38 Saenger, W., Hunter, W.N. and Kennard, O. (1986) *Nature*, **324**, 385–388.
- 39 Westhof, E. (1987) *Int. J. Biol. Macromol.*, **9**, 186–192.
- 40 Liu, K., Cruzan, J.D. and Saykally, R.J. (1996) *Science*, **271**, 929–933.
- 41 Neidle, S., Berman, H. and Shieh, H.S. (1980) *Nature*, **288**, 129–133.
- 42 Lubini, P., Zurcher, W. and Egli, M. (1994) *Chem. Biol.*, **1**, 39–45.